

# Bayesian Network-based Mendelian Randomization for Variant Prioritization and Phenotypic Causal Inference

Jianle Sun<sup>1</sup>, Jie Zhou<sup>1</sup>, Yuqiao Gong<sup>1</sup>, Chongchen Pang<sup>1</sup>,  
Yanran Ma<sup>1</sup>, Jian Zhao<sup>2</sup>, Zhangsheng Yu<sup>1\*</sup>, Yue Zhang<sup>1\*</sup>

<sup>1\*</sup>Department of Bioinformatics and Biostatistics, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, 200240, China.

<sup>2</sup>Ministry of Education and Shanghai Key Laboratory of Children's Environmental Health, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, 1665 Kongjiang Road, Shanghai, 200092, China.

\*Corresponding author(s). E-mail(s): [yuzhangsheng@sjtu.edu.cn](mailto:yuzhangsheng@sjtu.edu.cn);  
[yue.zhang@sjtu.edu.cn](mailto:yue.zhang@sjtu.edu.cn);

Contributing authors: [sjl-2017@sjtu.edu.cn](mailto:sjl-2017@sjtu.edu.cn); [jie.zhou@sjtu.edu.cn](mailto:jie.zhou@sjtu.edu.cn);  
[gyq123@sjtu.edu.cn](mailto:gyq123@sjtu.edu.cn); [pangchongchen@sjtu.edu.cn](mailto:pangchongchen@sjtu.edu.cn); [qiyumyr@sjtu.edu.cn](mailto:qiyumyr@sjtu.edu.cn);  
[jzhao.epi@gmail.com](mailto:jzhao.epi@gmail.com);

## Abstract

Mendelian randomization is a powerful method for inferring causal relationships. However, obtaining suitable genetic instrumental variables is often challenging due to gene interaction, linkage, and pleiotropy. We propose Bayesian Network-based Mendelian Randomization (BNMR), a Bayesian causal learning and inference framework using individual-level data. BNMR employs the random graph forest, a series of Bayesian network structural learning processes, to prioritize candidate genetic variants and select appropriate instrumental variables, and then obtains a pleiotropy-robust estimate by incorporating a shrinkage prior in the Bayesian framework. Simulations demonstrate BNMR can efficiently reduce the false positive discoveries in variant selection, and outperforms existing MR methods in terms of accuracy and statistical power in effect estimation. With application to the UK Biobank, BNMR exhibits its capacity in handling modern genomic data, and reveals the causal relationships from hematological traits to blood pressures and psychiatric disorders. Its effectiveness in handling complex

genetic structures and modern genomic data highlight the potential to facilitate real-world evidence studies, making it a promising tool for advancing our understanding of causal mechanisms.

**Keywords:** Mendelian Randomization, Variant Prioritization, Genome-wide Association Study, Bayesian Network, Horizontal Pleiotropy, Linkage Disequilibrium

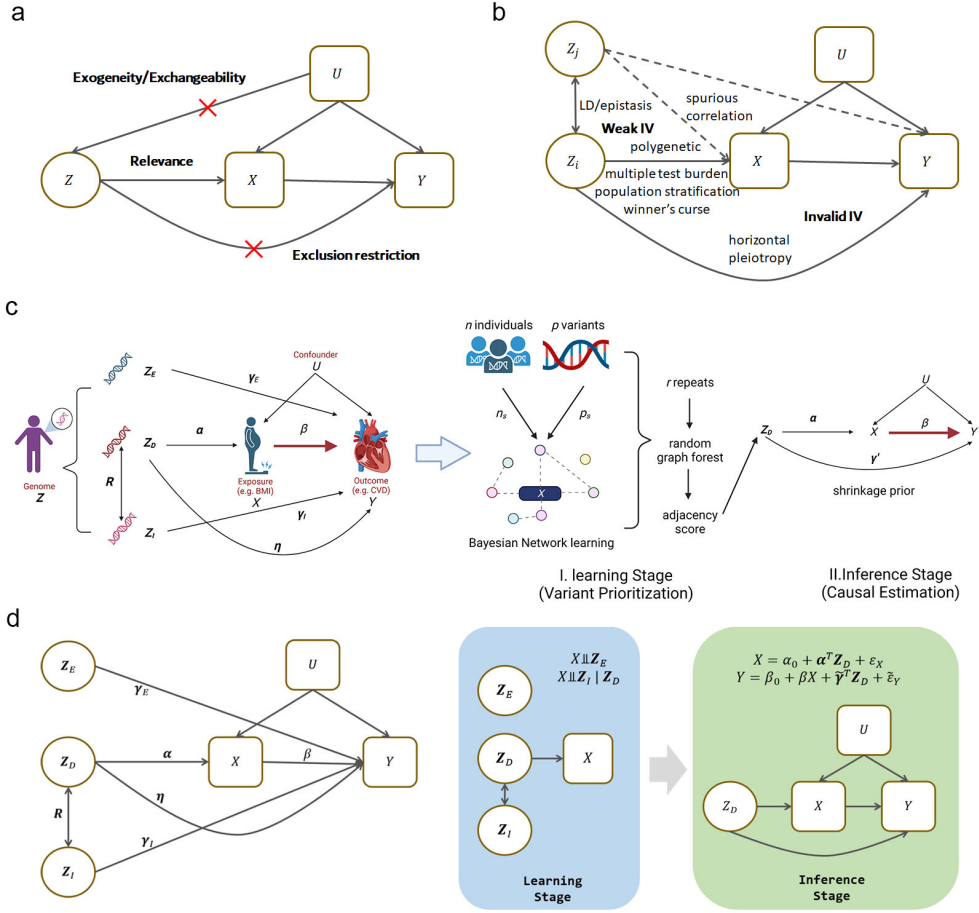
## Introduction

Identifying genuine causality is crucial to understanding physiological processes and discovering therapeutic targets, but it is also a tricky issue. Randomized controlled trials (RCTs) are usually regarded as the golden standard for causal inference but are restricted due to methodological, ethical, and economic concerns. Mendelian randomization (MR) is a promising approach to estimating causal effects using genetic variants as instrumental variables (IVs) (Sanderson et al, 2022). In general, MR analysis relies on three core assumptions: (i) relevance: a reliable correlation exists between the instrument and exposure; (ii) exogeneity or exchangeability: the instrument does not share common causes (confounders) with the exposure or outcome; and (iii) exclusion restriction: the instrument should affect the outcome only through the exposure (Fig 1a).

Unfortunately, the rigorous assumptions are often violated (Fig 1b), making it challenging to identify appropriate genetic instruments. First, although genome-wide association studies (GWAS) have identified numerous risk loci, in particular single nucleotide polymorphisms (SNPs), the effect on a polygenic complex trait is usually small, leading to weak-instrument bias (Davies et al, 2015). The multiple testing burdens, ‘winner’s curse’, linkage disequilibrium (LD), and population stratification increase the risk of false positive signals in GWAS (Tam et al, 2019). It can be improved by applying multiple instruments (Dudbridge, 2020), whereas correlated instruments will also lead to unstable estimates and introduce additional genetic confoundings when including non-causal variants (Gkatzionis et al, 2022). Proposed strategies such as LD stepwise pruning (Yang et al, 2012), principal components analysis (PCA) (Burgess et al, 2017), and penalization aim to extract a suitable number of independent instruments from a large set of correlated weak variants, but confront criticism on robustness (Gkatzionis et al, 2022).

Another problem is that many IVs are actually invalid due to horizontal pleiotropy (a variant affects the outcome via alternative pathways other than the exposure of interest). Gene interactions, such as LD and epistasis, can also violate exclusion restrictions analogous to pleiotropy. For individual-level data, lasso-type methods like sisVIVE (Kang et al, 2016) and post-adaptive Lasso (Windmeijer et al, 2019) help to control the influence of the pleiotropic effect. Recent approaches like TSHT (Guo et al, 2018) and CHIV (Windmeijer et al, 2021) mitigate pleiotropy by identifying valid instruments from candidate sets.

The above approaches relying on many plausible assumptions are tricky to model sophisticated real genetic patterns. Causal diagram model provides an alternative way



**Fig. 1** The overview of BNMR.

**a.** The three core assumptions of IV. **b.** The problems in current MR. Weak IVs are primarily due to the small individual contribution of a single locus to the trait and the low statistical power of GWAS, as well as the presence of linkage and interaction effects, leading to numerous false-positive discoveries. Invalid IVs are mainly caused by horizontal pleiotropy and linkage disequilibrium, which break the exclusion restriction assumption. **c.** The BNMR model. In the learning stage, we leverage the random graph forest to prioritize variants from a large interacting set and select variants with a true effect on the exposure as instruments. In the inference stage, we impose shrinkage prior on the Bayesian MR model to obtain a pleiotropy-robust estimate. **d.** Notations used in the BNMR model.

to represent the underlying causal relationships (Nogueira et al, 2022). With causal diagrams, machine learning techniques like the causal Bayesian network (BN) are currently applied to identify genetic interactions and causal variants (Lyu et al, 2021). They will also be a profitable complement to conventional MR (Howey et al, 2020; Amar et al, 2021).

In this paper, we propose a two-stage Bayesian Network-based Mendelian Randomization (BNMR) approach by integrating causal discovery and inference (Fig 1c). We aim to tackle correlated weak instruments in learning stage and cope with pleiotropy in inference stage. We prioritize effect variants and identify instruments from relevant variants with effects that are small and interacting using the random graph forest (RGF), comprised of a series of BN structure learning processes. Then we estimate the causal effects via a pleiotropy-robust MR approach, the Bayesian MR framework with a shrinkage prior to cope with potential horizontal pleiotropy (Berzuini et al, 2020). We demonstrate that BNMR is superior to conventional approaches in both instrument selection and effect estimation via simulations. With application to the UK BioBank, we examine causal pathways from hematological parameters to blood pressures and psychiatric disorders, bringing new biological insights.

## Methods

### Overview of the BNMR model

BNMR is a two-staged MR framework using individual-level data. In the learning stage, we propose RGF to select variants with reliable relevance from a large number of correlated weak instruments. We utilize BNs to characterize the complex conditional probability relationships and partition the variant set  $\mathcal{Z}$  into three subsets according to their relationships with the exposure of interest  $X$  (DIE partition),

$$\mathcal{Z} = \mathcal{Z}_D \cup \mathcal{Z}_I \cup \mathcal{Z}_E. \quad (1)$$

We use notations in calligraphic font to represent the variant set, bold font to represent the vector of genotypes, and *Italic capital letter* to represent single genotype. Variants in  $\mathcal{Z}_D$  directly affect the exposure, variants in  $\mathcal{Z}_I$  indirectly affect the exposure via gene interaction or linkage, i.e. variants in  $\mathcal{Z}_I$  and  $X$  are d-separated by variants in  $\mathcal{Z}_D$  ( $\mathbf{Z}_I \perp\!\!\!\perp X | \mathbf{Z}_D$ ), while variants in  $\mathcal{Z}_E$  do not affect the exposure ( $\mathbf{Z}_E \perp\!\!\!\perp X$ ). The three subsets are distinguished via BN under the causal Markov, faithfulness, and sufficiency assumptions (Nogueira et al, 2022), and only  $\mathcal{Z}_D$  can be parents of  $X$  in the causal graph.

In the inference stage, we model the potential horizontal pleiotropy explicitly. Since quantitative traits are determined by both genetic and environmental factors, assuming linearity and no interaction, we have

$$X = \alpha_0 + \boldsymbol{\alpha}^T \mathbf{Z}_D + \varepsilon_X, \quad (2)$$

and

$$Y = \gamma_0 + \boldsymbol{\gamma}_D^T \mathbf{Z}_D + \boldsymbol{\gamma}_I^T \mathbf{Z}_I + \boldsymbol{\gamma}_E^T \mathbf{Z}_E + \varepsilon_Y, \quad (3)$$

where  $\boldsymbol{\alpha}$  and  $\boldsymbol{\gamma}$  represent corresponding effect size on  $X$  and  $Y$ . Variants  $Z_D \in \mathcal{Z}_D$  affect the outcome  $Y$  through two different pathways: with the mediation of exposure  $X$  ( $\mathbf{Z}_D \xrightarrow{\boldsymbol{\alpha}} X \xrightarrow{\boldsymbol{\beta}} Y$ ), the causal pathway of interest, and via direct pathway or through other mediators other than  $X$  ( $\mathbf{Z}_D \xrightarrow{\boldsymbol{\eta}} Y$ ), known as (horizontal) pleiotropy

(Pingault et al, 2018). Under the assumption that both pathways are independent (the InSIDE assumption) (Pingault et al, 2018), we have

$$\boldsymbol{\gamma}_D = \beta \boldsymbol{\alpha} + \boldsymbol{\eta}, \quad (4)$$

where  $\beta$  is the causal effect of  $X$  on  $Y$ , while  $\boldsymbol{\eta}$  represents the pleiotropic effects. By introducing the correlation matrix  $\mathbf{R}$  between  $\mathbf{Z}_D$  and  $\mathbf{Z}_I$ ,

$$\mathbf{Z}_I = \mathbf{R}_0 + \mathbf{R} \mathbf{Z}_D + \boldsymbol{\varepsilon}_R, \quad (5)$$

we can rewrite the Eq. 2 and 3 as

$$\begin{aligned} X &= \alpha_0 + \boldsymbol{\alpha}^T \mathbf{Z}_D + \varepsilon_X, \\ Y &= \beta_0 + \beta X + \tilde{\boldsymbol{\gamma}}^T \mathbf{Z}_D + \tilde{\varepsilon}_Y, \end{aligned} \quad (6)$$

where  $\beta_0 = \boldsymbol{\gamma}_R^T \mathbf{R}_0 + \boldsymbol{\gamma}_E^T \mathbf{Z}_E - \beta \alpha_0$ ,  $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\eta} + \mathbf{R} \boldsymbol{\gamma}_I$ , and  $\tilde{\varepsilon}_Y = \boldsymbol{\gamma}_I^T \boldsymbol{\varepsilon}_R + \varepsilon_Y - \beta \varepsilon_X$ .  $\varepsilon_X$  and  $\tilde{\varepsilon}_Y$  are correlated, but are both independent with  $\mathbf{Z}_D$ . Actually, only a subset of variants in  $\mathbf{Z}_D$  need to be included as IVs, and consequently, variants in  $\mathbf{Z}_D$  are required to be identified with high precision. We then impose a shrinkage prior on nuisance parameters  $\tilde{\boldsymbol{\gamma}}$  to make  $\beta$  identifiable (Berzuini et al, 2020). Details on the derivation can be found in Supplementary Note SN1.

## BN structure learning in the random graph forest

To reduce the computational complexity of structure learning and assess confidence of each edge, we propose RGF, inspired by the random forest. In RGF,  $r$  sub-graphs are created using bootstrapping or subsampling, in which  $n_s$  of  $n$  individuals and  $p_s$  of  $p$  variants are sampled in each sub-graph. Consequently, we boil down the process of DIE partitioning to the structure learning of a series of causal BNs.

Since variants are reasons of traits naturally, we can simplify structure learning to graph skeleton determination. We identify  $\mathbf{Z}_D$  by scanning the variants directly adjacent to the exposure in each graph and calculating the adjacency score (the frequency of the  $Z - X$  edge presence in all sub-graphs) for each variant, which is the confidence of the variant-exposure relevance in the average causal graph. Variants with higher adjacency scores are at a higher confidence level to be identified as  $\mathbf{Z}_D \in \mathbf{Z}_D$ . We can select a specified number of lead variants or variants with scores higher than a given threshold  $\frac{\alpha^* p_s}{p}$  as IVs.

Various algorithms are proposed for BN structure learning. Scored-based approaches ascertain the optimal network by exhaustively or heuristically exploring candidate graphs and maximizing the network score, while constraint-based approaches leverage a sequence of conditional independence tests to establish the edge constraints between nodes and subsequently refine the directions (Nogueira et al, 2022). We implement score-based approaches including Hill-Climbing (hc) and Tabu Search (tabu), constrained-based approaches including stable PC (pc.stable), Incremental Association (iamb), and Grow-Shrink (gs), as well as hybrid learning methods

including Max-Min Hill-Climbing (mmhc) and Restricted Maximization (rsmx2). All these methods are implemented using the R package bnlearn.

## Bayesian MR estimation with a shrinkage prior

We specify model (6) under the Bayesian framework. The total error term  $\varepsilon_X$  and  $\tilde{\varepsilon}_Y$  can be decomposed into a confounding-related term  $\delta$  and a completely random term  $\sigma$ , i.e.,  $\mathbb{E}(\varepsilon_X) = \mathbb{E}(\tilde{\varepsilon}_Y) = 0$ ,  $\text{Var}(\varepsilon_X) = \delta_1^2 + \sigma_1^2$ , and  $\text{Var}(\tilde{\varepsilon}_Y) = \delta_2^2 + \sigma_2^2$ . Assuming that the two completely random terms are uncorrelated, we have  $\text{Cov}(\varepsilon_X, \tilde{\varepsilon}_Y) = \delta_1\delta_2$ .

We only need to select a subset of  $\mathcal{Z}_D$  as instruments, and have the Bayesian MR model (Berzuini et al, 2020)

$$\begin{aligned} X|Z, U &\sim \mathcal{N}(\alpha_0 + \sum_{j=1}^J \alpha_j Z_j + \delta_1 U, \sigma_1^2) \\ Y|X, Z, U &\sim \mathcal{N}(\beta_0 + \beta X + \sum_{j=1}^J \gamma_j Z_j + \delta_2 U, \sigma_2^2), \\ U &\sim \mathcal{N}(0, 1). \end{aligned} \tag{7}$$

where  $Z_j \in \mathcal{Z}_D$ . To make causal parameter  $\beta$  identifiable, we assume that not all IVs selected take pleiotropic effects (i.e., some components of  $\boldsymbol{\gamma}$  are zero) and impose a shrinkage prior on  $\boldsymbol{\gamma}$  under the Bayesian framework (Berzuini et al, 2020). The Bayesian estimation is conducted using Markov Chain Monte-Carlo (MCMC) with Rstan and PyMC. The first half of the iteration is used for burn-in, and the second half is used for sampling.

BNMR can extend to binary outcomes by modifying the Eq. 7 to logistic regressions, i.e.,

$$Y|X, Z, U \sim \text{Bernoulli}(\text{logit}^{-1}(\beta_0 + \beta X + \sum_{j=1}^J \gamma_j Z_j + \delta_2 U)). \tag{8}$$

We compare estimates of BNMR with other IV selection and MR estimation approaches. We implement PCA with R package stats and penalized regressions with R package glmnet, where 10-fold cross validation is used to determine the best value of  $\lambda$ . Compared methods are implemented with the R package AER, ivmodel, Mendelian-Randomization, sisVIVE, cause, R2BGLiMS, and CIIV. We implement BNMR as an R package, with source codes available at <https://github.com/sjl-sjtu/bnmr>.

## Simulations

We use both simulated and real genomics from UK Biobank in simulations. For simulated genomics,  $k$  independent loci sampled from multinomial distributions, whose genotype frequencies satisfy the Hardy-Weinberg equilibrium (HWE), with the effect allele frequency  $\pi$  from  $U(0.05, 0.95)$ .  $m$  correlated loci for each locus are simulated according to LD squared correlation coefficient ( $r^2$ ) (Pritchard and Przeworski, 2001)

that sampled from  $U(0.05, 0.95)$ , and genomics with  $p = k(m + 1)$  loci are synthesized. Real genomic data used to simulate phenotypes are derived from variants on chromosomes 10, 17 and 22 in the European ancestry population of UK Biobank.

Phenotypes are generated from linear model

$$X = \alpha_0 + \sum_j \alpha_j G_{X,j} + \delta_x U + \varepsilon_x, \varepsilon_x \sim \mathcal{N}(0, \sigma_x^2), \quad (9)$$

and

$$Y = \beta_0 + \beta X + \sum_k \gamma_k G_{Y,k} + \delta_y U + \varepsilon_y, \varepsilon_y \sim \mathcal{N}(0, \sigma_y^2), \quad (10)$$

with causal effect  $\beta = 0.5$ . Confounder  $U$  is generated from the standard Gaussian distribution  $\mathcal{N}(0, 1)$ , with coefficients  $\delta_x = \delta_y = 1$ . Variants affecting  $X$  are randomly selected from the simulated genome, with effect size  $\alpha_j \sim 0.1 + |\mathcal{N}(0, 0.05^2)|$ . Variants affecting  $Y$  are either non-pleiotropic or pleiotropic (those that also affect  $X$ ). For non-pleiotropic variants  $\gamma_k \sim \mathcal{N}(0.1, 0.05^2)$ , and for pleiotropic variants  $\gamma_k \sim \mathcal{N}(\mu_\gamma, 0.05^2)$ , with  $\mu_\gamma = 0$  (balanced pleiotropy) or  $\mu_\gamma = 0.05$  (directional pleiotropy). We utilize 50 replicates for each scenario and report the average.

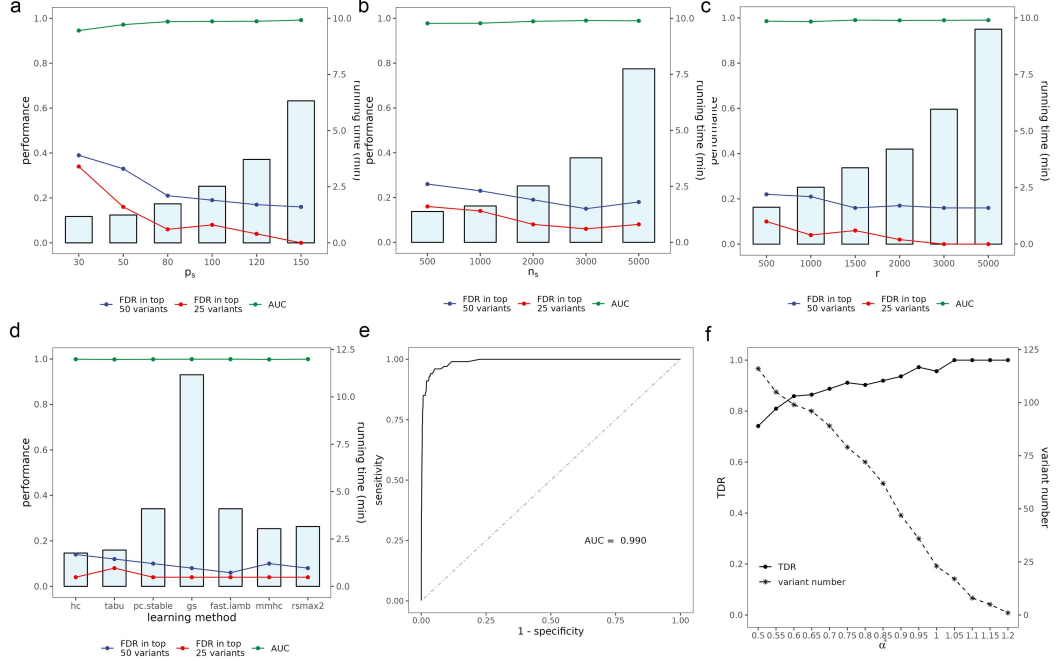
For scenarios with non-additive genetic effects, we simulate phenotypes using more complicated polygenetic models with simple multiplicative effects, interactive multiplicative effects, and interactive threshold effects (Marchini et al, 2005). Details can be found at Supplementary Note SN3.

## Results

### **BNMR can efficiently identify effect variants from numerous weak, interacting variants with high precision in learning stage.**

We first compare the fine-mapping performance of RGF with different hyperparameters and structure learning algorithms in simulated datasets (Fig 2). RGF exhibits a lower false discovery rate (FDR) and a higher AUC, with increasing subsample size and numbers of subsamples, though this improvement is accompanied by an increase in time consumption. Constrained-based approaches yield lower FDR, while the score-based approaches are superior in speed. As the selection threshold increases, the number of identified variants diminishes with increased precision.

Compared to the conventional association test (linear regression), RGF, LD step-wise pruning, and penalized regressions (especially lasso and elastic net) can all reduce FDR, while the RGF achieves the highest precision, performing as a effective tool in prioritizing candidate effect variants and identifying true effect variants ( $\mathcal{Z}_D$ ) (Fig 3). Before employing these variable selection strategies, we conduct pre-filtering to reduce the number of candidate variants. A more strict  $P$  threshold before RGF increases the precision in top variants but may cause more false discoveries. A threshold of around  $\frac{1}{p}$  to  $\frac{0.01}{p}$  may be proper. Besides, adjacency score can be regarded as the confidence that the variant belongs to  $\mathcal{Z}_D$ , and thus, it is also an assessment of instrument strength. The correlation between the adjacency score and the commonly used F statistics (Fig



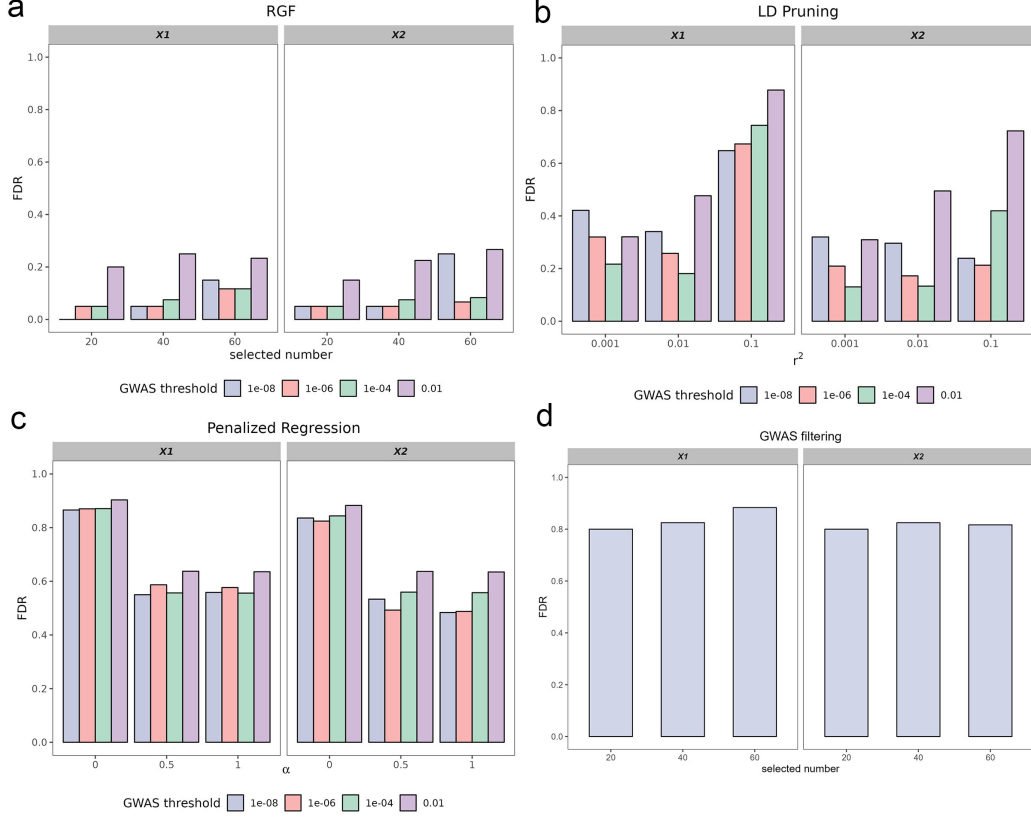
**Fig. 2** The performance of the RGF with different hyperparameters and BN structure learning methods in simulations.

**a.** The performance and time consumption of RGF with different numbers of subsampling variants using the Hill-Climbing (hc) algorithm. **b.** The performance and time consumption of RGF with different numbers of subsampling individuals using the hc algorithm. **c.** The performance and time consumption of RGF with different numbers of subsamples. **d.** The performance and time consumption of RGF with different BN structure learning methods. We evaluate score-based approaches including hc and tabu, constrained-based approaches including stable PC (pc.stable), Incremental Association (iamb), and Grow-Shrink (gs), as well as hybrid learning methods including Max-Min Hill-Climbing (mmhc) and Restricted Maximization (rsmx2). The lines show the corresponding FDR and AUC, while the gray bars display the changes in consumed time (min). **e.** The ROC curve for RGF ( $n_s = 2000$ ,  $p_s = 120$ ,  $r = 1000$ ). **f.** The relationships among selection threshold ( $\alpha^*$ ), number of selected variants, and FDR. Simulated data size:  $n = 5000$ ,  $p = 2000$ , with 100 true effect variants for the exposure. FDR: false discovery rate. AUC: area under the receiver operating characteristic (ROC) curve.

S2) indicate that RGF is capable of choosing instruments reliable relevance to reduce weak-instrument bias.

GWAS becomes tricky when dealing with non-additive genetic effects (Tam et al, 2019). We simulate phenotypes from the combination of three interacting polygenetic models, including simple multiplicative effects, interactive multiplicative effects, and interactive threshold effects (Supplementary Note SN3) (Marchini et al, 2005). The results show that RGF performs well in settling the puzzle of gene interaction and epistasis (Fig S4).

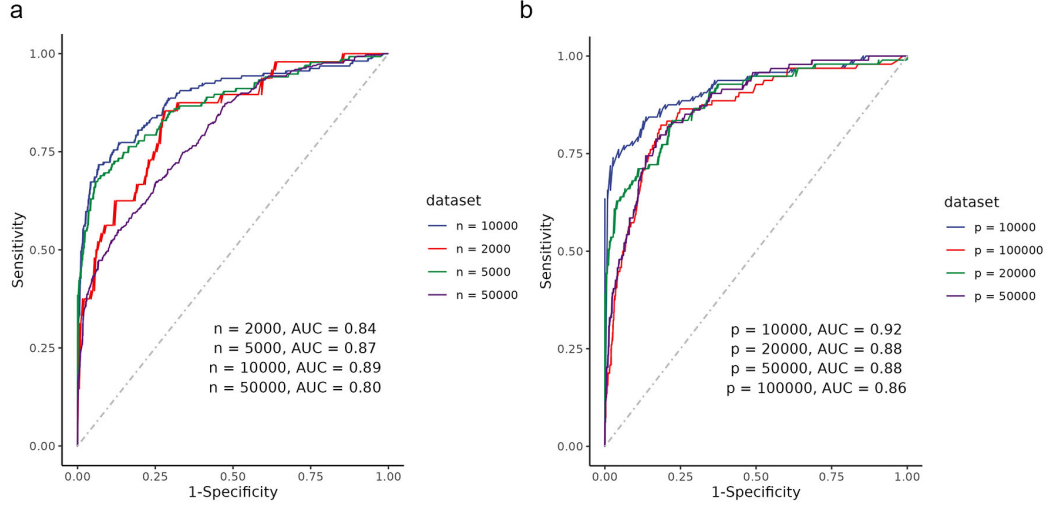




**Fig. 3** False discovery rate (FDR) of different IV selection strategies in simulations.

**a.** FDR of RGF. **b.** FDR of LD stepwise pruning.  $r^2$ : correlation thresholds for LD pruning. **c.** FDR of penalized regressions.  $\alpha$  is set to be 1.0 (Lasso), 0.5 (elastic net), and 0 (ridge regression), with penalty factor  $\lambda$  determined by 10-fold cross validation. **d.** FDR of GWAS lead SNPs. Variants are pre-filtered according to specific GWAS  $P$ -value thresholds before LD pruning, penalized regressions, or RGF. Simulated data size:  $n = 10000$ ,  $p = 10000$ , with 300 true effect variants for each trait. The environmental variance ( $\sigma_x^2$ ) for  $X_2$  is imposed to be twice as much as that of  $X_1$  to represent traits with lower heritability. RGF is conducted with  $n_s = 2000$ ,  $p_s = 150$ ,  $r = 5000$  using hc algorithm.

Since genotypes are notoriously difficult to simulate, we also generate phenotypes using the same procedures (Eq. 11) but based on real genotype data from the UK BioBank. Using the synthetic datasets, we demonstrate the adaptation of RGF to different scales of genomes (Fig 4). RGF is capable of handling the complex structure of real genetic data and exhibits good adaptability to datasets with different scales.

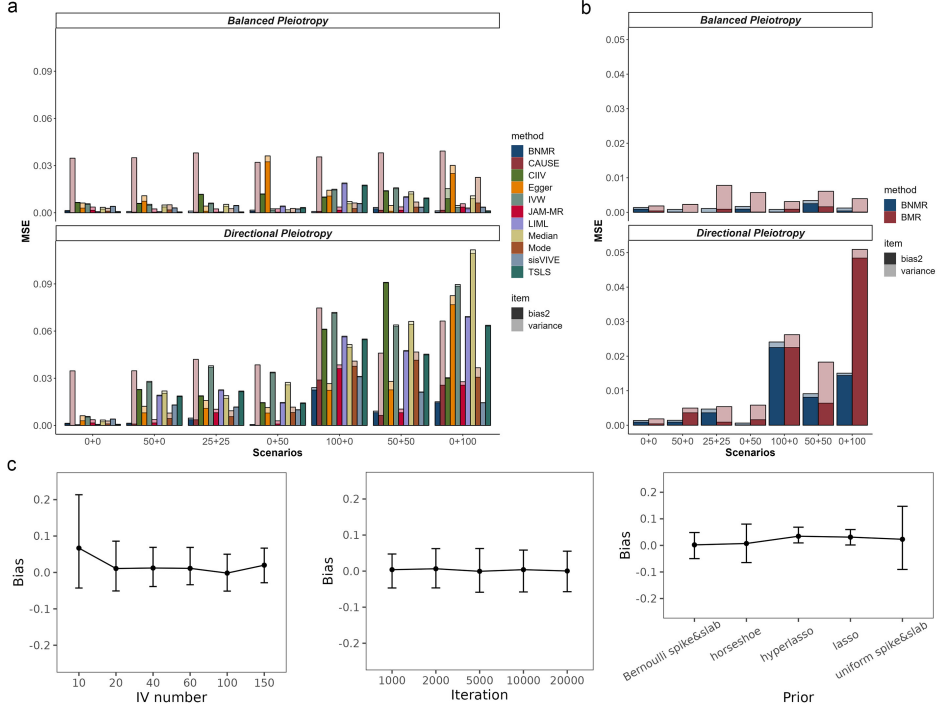


**Fig. 4** Performance of BNMR on simulated phenotypes based on real genotypes from the UK Biobank.

**a.** The ROC curves for real genotype data with different sample size ( $n$ ). Here  $p$  is fixed at 20000. **b.** The ROC curves for real genotype data with different genome size ( $p$ ). Here  $n$  is fixed at 5000.

### BNMR can effectively reduce mean square error of estimates, enhance statistical power, and is robust to horizontal pleiotropy in inference stage.

We first compare the performance of BNMR to other existing MR approaches, including two-stage least square (TSLS), limited information maximum likelihood ratio (LIML) (Boehm and Zhou, 2022), inverse-variance weighted (IVW) (Burgess et al, 2013), MR-Egger (Bowden et al, 2015), weighted median (Bowden et al, 2016), weighted mode (Hartwig et al, 2017), sisVIVE (Kang et al, 2016), JAM-MR (Gkatzionis et al, 2021), CAUSE (Morrison et al, 2020), and CIIV (Windmeijer et al, 2021) (Fig 5a). We include two types of pleiotropic loci in our simulation: true level pleiotropy loci that simultaneously affect exposure and outcome, and correlated pleiotropy resulting from loci in linkage disequilibrium that independently affect exposure and outcome. For each scenario, we examine the performance of various methods under settings where the expected average pleiotropic effect of all loci was either 0 (balanced pleiotropy) or non-zero (directional pleiotropy). Oracle approaches like TSLS and LIML fail to cope with scenarios with complex pleiotropy. Some pleiotropy-robust summary data-based methods, like CAUSE, reduce the bias effectively while confronting an inflated estimation variance when applying to one-sample studies. On the contrary, BNMR outperforms in statistical power due to a smaller variance of estimates. BNMR is superior to the existing approaches in terms of mean squared error (MSE) in spite of somewhat weaker performance for directional pleiotropy.



**Fig. 5** Performance on causal effect estimation in simulations.

**a.** Mean square errors (MSE) of different MR approaches. Robust regression and penalized weights are adopted in IVW and MR-Egger. IVs used in TSLS, LIML, IVW, MR-Egger, weighted Median and Mode, and CIIV are obtained through LD stepwise pruning. IVs used in sisVIVE, JAM-MR, CAUSE, and BNMR are selected using their own filtering procedures from GWAS lead SNPs. Standard errors (SEs) of sisVIVE are calculated via bootstrapping, of CAUSE are transformed from confidence intervals, of BNMR are calculated from MCMC posterior sampling. The scenarios are represented as “the number of true pleiotropic variants (variants with direct effects on both  $X$  and  $Y$ ) + the number of correlated pleiotropic variants (variants with causal effects on  $Y$  are correlated to variants with causal effects on  $X$ )”. For each scenario, the pleiotropic effects are simulated to be either balanced (mean pleiotropic effect  $\mu_Y = 0$ ) or directional (mean pleiotropic effect  $\mu_Y = 0.05$ ). Each bar represents the average MSE of the estimation for the respective method in that scenario. The darker section at the bottom represents the squared bias, while the lighter section at the top represents the variance. **b.** MSE of Bayesian MR using IVs from BN (BNMR) and using the same number of IVs from GWAS lead SNPs (BMR). **c.** The average bias and error bars under different IV numbers (left), iterations (middle), and shrinkage priors (right).

To show the bonus BN brings to the conventional Bayesian MR, we then evaluate the improvement in Bayesian MR estimation by using IVs obtained from BN (BNMR) compared to using GWAS lead SNPs directly as IVs (BMR). A noticeable reduction in MSE can be observed (Fig 5b), especially when there is presence of unbalanced (directional) pleiotropy. In the case of balanced pleiotropy, this enhancement primarily manifests in the attenuation of estimated variance, thereby yielding augmented statistical power.

We also conduct sensitivity analysis on the instrument numbers and iterations, and different shrinkage priors (Van Erp et al, 2019) (Fig 5, Supplementary Note SN4-SN5). Bias increases when there are too many or too few instruments. BNMR estimates are not sensitive to priors in general, despite the fact that uniform spike & slab prior is a bit more inefficient than the others based on the error bars (Fig 5c). The Bayesian Lasso prior shows the fastest sampling speed, while the horseshoe prior, although slightly less efficient, is superior in the performance of convergence due to the lowest Rhat (Table S4 & Fig S5).

## **BNMR with large-scale biobank-level data vindicates causality from erythrocyte-related traits to blood pressures.**

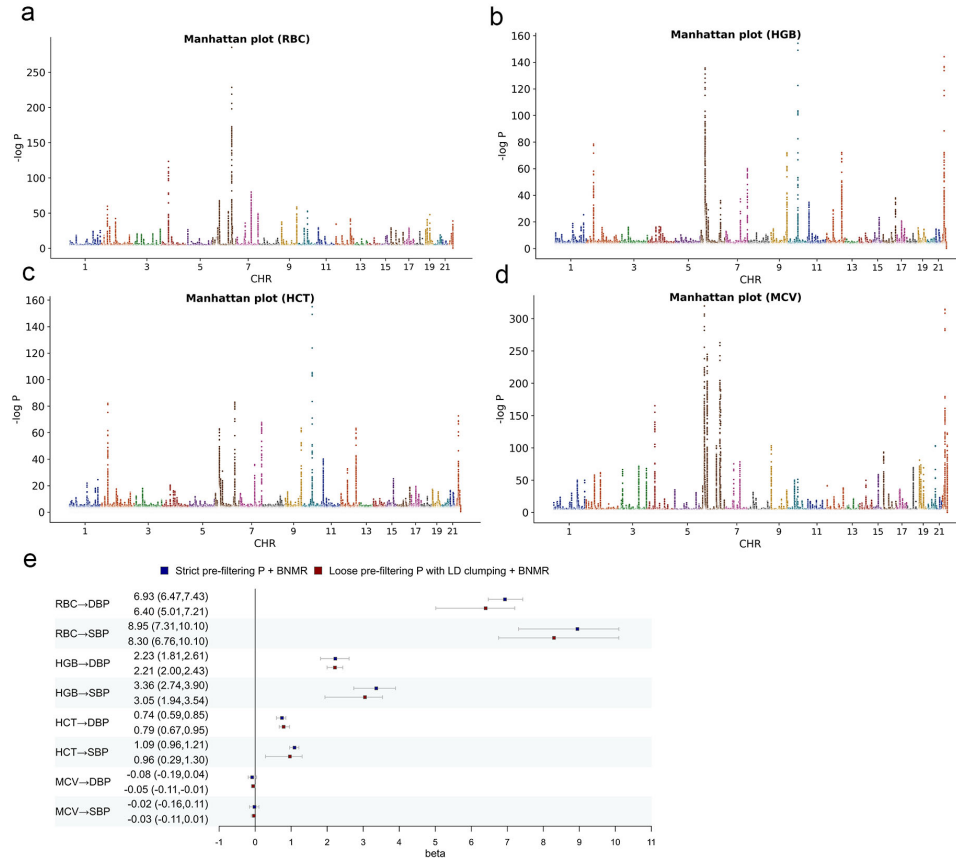
To highlight the practical significance and applicability of our method on extensive modern genetic datasets, we provide illustrative examples of two real-world studies featuring both continuous and binary outcomes, utilizing data sourced from the UK BioBank.

Hematological indices usually vary in a variety of physiological processes and are potential indicators for related disorders. Correlational studies have proposed that erythrocyte-related characteristics, including red blood cell count (RBC), hemoglobin concentration (HGB), hematocrit (HCT), the proportion of RBCs to the plasma, and mean red cell volume (MCV) (Enawgaw et al, 2017), are in strong correlation with systolic and diastolic blood pressures (SBP and DBP), and abnormalities of erythrocytes might be indicators of some cardiovascular and cerebrovascular diseases such as hypertension (Tsuda, 2020).

To examine the causal effects of erythrocyte parameters on blood pressures, we involve 246,659 participants of Caucasian ancestry, self-reported as free from hypertension or other cardiovascular diseases (UK Biobank Non-cancer Illness Coding 1065-1094), and with available blood routine measurements at the time of enrollment. enome quality control is conducted using PLINK 2.0, with thresholds for the SNP missing rate, minor allele frequency (MAF), and HWE test are 0.05, 0.01, and 1e-6. Fast posterior sampling with the large dataset is conducted with the Python packages PyMC and JAX. To increase power, we conducted preliminary GWAS filtering using summary statistics from a different dataset of the same ethnic group but with distinct sample (Astle et al, 2016).

Two pre-filtering strategies for BNMR (Fig 6e) consistently show that RBC, HGB, and HCT show significant positive effects on both DBP and SBP, and the effect magnitude is larger on SBP than on DBP. Whereas MCV shows a non-significant negative effect on blood pressure instead. Alternative approaches use top GWAS significant SNPs after LD clumping as instruments, and the varied and even conflicting results (Table S5) remind us of the importance of MR methodology. MR-Egger test shows that all causal relationships are not significant. On the other hand, TSLS indicates that RBC, HGB, and HCT have a significant effect on DBP but not on SBP, while CIIV estimates a positive effect on DBP and a negative effect on SBP.

The underlying mechanisms may relate to blood viscosity. Higher RBC, HGB, and HCT mean an increase in blood viscosity and peripheral resistance to blood flow, resulting in hypertension (Enawgaw et al, 2017). Besides, erythrocytes and hemoglobin



**Fig. 6** Causal relationships from erythrocyte-related traits to blood pressures. Manhattan plots for RBC (a), HGB (b), HCT (c), and MCV (d), where the vertical coordinate shows the negative logarithm of the GWAS association test  $P$ -values for each locus, and a forest plot of the causal estimations (e). The units of RBC, HGB, HCT, MCV, and blood pressure are million/mm<sup>3</sup>, g/dl, %, fL, and mmHg. We adopt two different pre-filtering strategies: one uses a more strict GWAS  $P$  threshold ( $1e-20$ ), the other uses a looser  $P$  threshold ( $5e-8$ ) and then conducts LD clumping (threshold: window = 10000kb,  $r^2=0.01$ , MAF=0.01). The variant loci obtained from pre-filtering are then further selected via RGF ( $n_s = 4000$ ,  $p_s = 150$ ,  $r = 5000$ ) to identify 20 instruments for each exposure, shown in Supplementary Note SN6, and sample 4 chains with 5000 iterations per chain in MCMC.

also influence nitric oxide bioavailability, a crucial signal in the regulation of vessel psychology such as vasodilatation, thrombosis inhibition, and vessel formation (Helms et al, 2018). Although the molecular mechanisms still remain to be uncovered, the findings indicate that those hematological indices may be not only indicators but potential therapeutic targets for hypertension.

## BNMR indicates that increased lymphocytes contribute to the risk of schizophrenia.

The neuro-immune interaction has been an appealing topic in recent years. Widespread bidirectional circuits exist between the two systems. The nervous system regulates immune activity and cytokine balance via the direct connection of sympathetic and parasympathetic nerves, and some neurotransmitters and neuroendocrine hormones can also serve as immunomodulators. Meanwhile, the immune system participates in the elimination and plasticity of synapses during development and modulates brain activity as well (Dantzer, 2018). Immune-related hematological biomarkers provide a new insight into the pathological mechanisms of many psychiatric disorders. For instance, a recent study reports lymphocyte count is associated with psychiatric disorders like schizophrenia (SCZ) and post-traumatic stress disorder (PTSD) (Reay et al, 2022).

We leverage disease records from UK Biobank and construct a case-control study by randomly selecting healthy individuals of the same ethnicity to assess whether lymphocyte count will causally affect SCZ and PTSD. 411 subjects with SCZ and 253 subjects with PTSD are included, and we sample 1000 subjects without psychiatric disorders to balance the sample. All participants are of Caucasian ancestry.

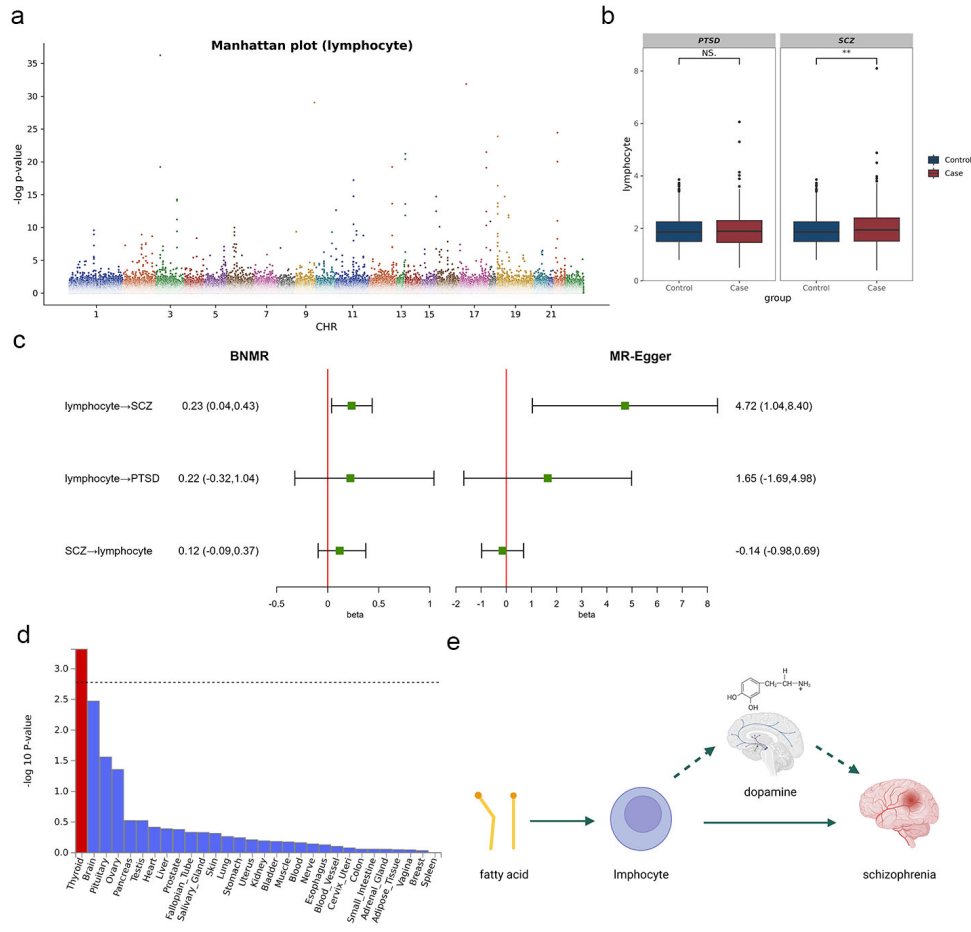
We leverage previous GWAS summary statistics for lymphocyte counts (Astle et al, 2016) and SCZ (Jiang et al, 2021) to pre-filter candidate variants before RGF.

Results from both BNMR and MR-Egger indicate that lymphocyte count has a positive effect on SCZ but have no significant effect on PTSD (Fig 7c). The reciprocal MR analysis supports the causal direction from lymphocyte count to SCZ.

We further conduct gene mapping and functional annotation by FUMA (Watanabe et al, 2017) using the top 1500 variants identified in RGF, which maps 303 lymphocyte-related and 36 SCZ-related protein-coding genes (Supplementary Note SN8). MAGMA tissue expression analysis reveals an enrichment of SCZ-related genes in the thyroid (Fig 7d), the organ secreting hormones as essential immune modulators. The secretion is regulated by dopamine, an essential neuro-immune-transmitter that has been a focus of SCZ aetiology research for decades because a deficit in dopamine release is reported in SCZ patients (Weinstein et al, 2017). SCZ is reported to be associated with immune system dysfunction, manifesting in abnormal blood immune parameters (Miller et al, 2013). Meanwhile, lymphocytes participate in the synthesis of dopamine and its metabolites, and abnormalities in dopamine receptor expression in lymphocytes are observed in SCZ patients (Penedo et al, 2021).

Previous MR research identified the protective effects of long-chain  $\omega$ -3 and  $\omega$ -6 fatty acids on SCZ (Jones et al, 2021). Meanwhile, substantial evidence from both population research and cellular experiments demonstrates that decreases in fatty acids cause lymphocyte expansion (Kelley, 2001). We can therefore conjecture a possible causal pathway: plasma fatty acids  $\rightarrow$  lymphocyte  $\rightarrow$  dopamine  $\rightarrow$  SCZ (Fig 7e).

Though there are currently no validated biomarkers for the diagnosis and prognosis of SCZ, our analysis supports the idea of modulating immune functions as an intervention for SCZ. Since lymphocytes are further downstream than fatty acids in the conceived pathway, modulation of lymphocytes is likely to cause more direct protective effects and fewer side effects. However, the results should still be interpreted with



**Fig. 7** The relationships between lymphocyte count and psychiatric disorders. **a.** Manhattan plots for lymphocyte count. The vertical coordinate shows the negative logarithm of the GWAS association test  $P$ -values for each locus. **b.** Differences in lymphocyte count between the case and control groups. The significance level is calculated using the T test. **c.** Forest plot of the causal estimations of BNMR and MR-Egger. The unit of lymphocyte count is billion cells per liter. For BNMR, we select 20 instruments for each exposure via RGF ( $n_s = 4000$ ,  $p_s = 150$ ,  $r = 5000$ ), shown in Supplementary Note SN6, and sample 4 chains with 5000 iterations per chain in MCMC. For MR-Egger, GWAS lead variants after LD clumping are selected as instruments. **d.** MAGMA tissue expression analysis for mapped SCZ-related genes in GTEx v8 30 general tissue types. **e.** A potential causal pathway of SCZ: fatty acids regulate the lymphocyte counts, and lymphocytes contribute to the development of SCZ by participating in dopamine metabolism.

caution due to recipient inclusion and sample size, population heterogeneity, and other potential confounding factors. Noncollapsibility of the logistic model may damage the estimation of binary responses (Schuster et al, 2021). Collaboration with evidence by

means of triangulation (Lawlor et al, 2016) is vital to drawing a solid and reliable conclusion.

## Discussion

Causality is challenging to identify in observational studies due to unmeasurable confounders. The introduction of genetic instruments in MR makes it possible to estimate causal effect in the presence of unobserved confounders, making MR increasingly appealing in real-world studies.

Tackling imperfect IVs has always been a tricky issue in MR. We propose BNMR to address the challenges by leveraging machine learning techniques and integrating causal discovery and inference. We use RGF to reduce FDR and improve statistical power when selecting instruments with true effects from numerous correlated weak variants due to polygenicity, epistasis, and LD. Then we control horizontal pleiotropy by imposing a shrinkage prior on the Bayesian MR. We illustrate an improved accuracy and efficiency in simulations.

To guarantee the faithfulness and sufficiency of causal diagrams, we impose constraints in RGF that limit the nodes in the graph to only include genetic loci and a single exposure. We tend not to involve multiple traits in a causal graph because the common causes of those traits may not be observed. Another advantage is that the criteria ‘not d-separated from exposure by other variants’ can be simply expressed as ‘adjacent to exposure’ in this scenario, which is convenient for DIE partition and IV selection.

Bayesian estimation with imposed shrinkage priors is conceptually similar to regularization in the traditional model but with some obvious advantages, like simultaneously estimated penalty parameters, easily obtained credible intervals, and intuitive interpretation. Additionally, domain-specific knowledge can be included as an informative prior. BNMR are not sensitive to priors, though we recommend horseshoe prior for better convergence performance if no additional information is accessible.

Although large-scale biobanks containing genotypes and phenotypes are now available, an increasing number of studies tend to report summary association statistics instead due to concerns on privacy and security. Bayesian meta-analysis is applied to assess pooled genetic relevance (Sun et al, 2022). Recent work has started to focus on learning causal diagrams with summary data (Zhang et al, 2017), while arduous task still remains.

The model confronts computational challenges in BN learning and MCMC sampling, especially with increasing numbers of samples and variables. BN structural learning is an NP-hard problem. We leverage bagging techniques in ensemble learning and propose the RGF to split the whole genetic pattern into a series of subgraphs. The number of candidate variants is restricted via pre-filtering by GWAS association tests before RGF, since the removal of variants with low correlations will not influence the network structure severely due to the modularity of the causal diagram. To achieve a balance between sufficiently high precision and acceptable time consumption, we suggest conducting pre-filtering using a GWAS  $P$  threshold of approximately  $\frac{1}{p}$  to  $\frac{0.01}{p}$  and setting the value of  $p_s r$  to be at least 100 times greater than the number



of variants in RGF to ensure adequate sampling for each variant. We endorse the use of at least 4 chains and around 5000 iterations in MCMC. For large-scale datasets, consolidation of posterior sampling in subsamples may be feasible.

In summary, BNMR is a practical model to prioritize and select proper instruments from massive, interacting, and weak variants and obtain pleiotropy-robust causal effect estimates. With accumulated genomic data available, BNMR will contribute to revealing more causal relationships and discovering potential therapeutic targets with real-world evidence.

## **Statements & Declarations**

### **Funding**

The research is supported by the National Natural Science Foundation of China (11901387 for YZ and 12171318 for ZY).

### **Competing interests**

No competing interest exists.

### **Author contributions**

JS and YZ designed the study. JS performed the research, analyzed the data, and wrote the original manuscript. YM participated in the real-world data management. ZY and YZ supervised the research. JS, JZ, YG, CP, YM, JZ, ZY, and YZ discussed and revised the manuscript.

### **Data and code availability**

The GWAS summary statistics are accessed from GWAS Catalog (<https://www.ebi.ac.uk/gwas/downloads/summary-statistics>). The real-world data underlying this article are accessed from UK Biobank (<https://www.ukbiobank.ac.uk/>). We implement the BNMR framework as an R package, and all the codes are available at <https://github.com/sjl-sjtu/bnmr>.

### **Ethical approval**

UK Biobank has approval from the North West Multi-centre Research Ethics Committee (21/NW/0157). The real-world studies have been conducted using the UK Biobank Resource under Application Number 100014.

### **Consent to participate**

Not applicable.

### **Consent to publish**

Not applicable.

## Supplementary data

Supplementary Notes, Tables and Figures.

## References

- Amar D, Sinnott-Armstrong N, Ashley EA, et al (2021) Graphical analysis for phenome-wide causal discovery in genotyped population-scale biobanks. *Nature Communications* 12(1):1–11
- Astle WJ, Elding H, Jiang T, et al (2016) The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* 167(5):1415–1429
- Berzuini C, Guo H, Burgess S, et al (2020) A Bayesian approach to Mendelian randomization with multiple pleiotropic variants. *Biostatistics* 21(1):86–101
- Boehm FJ, Zhou X (2022) Statistical methods for Mendelian randomization in genome-wide association studies: A review. *Computational and Structural Biotechnology Journal* 20:2338–2351
- Bowden J, Davey Smith G, Burgess S (2015) Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* 44(2):512–525
- Bowden J, Davey Smith G, Haycock PC, et al (2016) Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genetic Epidemiology* 40(4):304–314
- Burgess S, Butterworth A, Thompson SG (2013) Mendelian randomization analysis with multiple genetic variants using summarized data. *Genetic Epidemiology* 37(7):658–665
- Burgess S, Zuber V, Valdes-Marquez E, et al (2017) Mendelian randomization with fine-mapped genetic data: choosing from large numbers of correlated instrumental variables. *Genetic Epidemiology* 41(8):714–725
- Dantzer R (2018) Neuroimmune interactions: from the brain to the immune system and vice versa. *Physiological Reviews* 98(1):477–504
- Davies NM, von Hinke Kessler Scholder S, Farbmacher H, et al (2015) The many weak instruments problem and Mendelian randomization. *Statistics in Medicine* 34(3):454–468
- Dudbridge F (2020) Polygenic Mendelian randomization. *Cold Spring Harbor Perspectives in Medicine* p a039586
- Enawgaw B, Adane N, Terefe B, et al (2017) A comparative cross-sectional study of some hematological parameters of hypertensive and normotensive individuals at the

- university of gondar hospital, Northwest Ethiopia. *BMC Hematology* 17(1):1–7
- Gkatzionis A, Burgess S, Conti DV, et al (2021) Bayesian variable selection with a pleiotropic loss function in Mendelian randomization. *Statistics in Medicine* 40(23):5025–5045
- Gkatzionis A, Burgess S, Newcombe PJ (2022) Statistical methods for cis-Mendelian randomization with two-sample summary-level data. *Genetic Epidemiology*
- Guo Z, Kang H, Tony Cai T, et al (2018) Confidence intervals for causal effects with invalid instruments by using two-stage hard thresholding with voting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 80(4):793–815
- Hartwig FP, Davey Smith G, Bowden J (2017) Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *International Journal of Epidemiology* 46(6):1985–1998
- Helms CC, Gladwin MT, Kim-Shapiro DB (2018) Erythrocytes and vascular function: oxygen and nitric oxide. *Frontiers in Physiology* 9:125
- Howey R, Shin SY, Relton C, et al (2020) Bayesian network analysis incorporating genetic anchors complements conventional Mendelian randomization approaches for exploratory analysis of causal relationships in complex data. *PLoS Genetics* 16(3):e1008198
- Jiang L, Zheng Z, Fang H, et al (2021) A generalized linear mixed model association tool for biobank-scale data. *Nature Genetics* 53(11):1616–1621
- Jones HJ, Borges MC, Carnegie R, et al (2021) Associations between plasma fatty acid concentrations and schizophrenia: a two-sample Mendelian randomisation study. *The Lancet Psychiatry* 8(12):1062–1070
- Kang H, Zhang A, Cai TT, et al (2016) Instrumental variables estimation with some invalid instruments and its application to Mendelian randomization. *Journal of the American statistical Association* 111(513):132–144
- Kelley DS (2001) Modulation of human immune and inflammatory responses by dietary fatty acids. *Nutrition* 17(7-8):669–673
- Lawlor DA, Tilling K, Davey Smith G (2016) Triangulation in aetiological epidemiology. *International Journal of Epidemiology* 45(6):1866–1886
- Lyu R, Sun J, Xu D, et al (2021) GESLM algorithm for detecting causal snps in GWAS with multiple phenotypes. *Briefings in Bioinformatics* 22(6):bbab276
- Marchini J, Donnelly P, Cardon LR (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nature Genetics* 37(4):413–417

- Miller BJ, Gassama B, Sebastian D, et al (2013) Meta-analysis of lymphocytes in schizophrenia: clinical status and antipsychotic effects. *Biological Psychiatry* 73(10):993–999
- Morrison J, Knoblauch N, Marcus JH, et al (2020) Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nature Genetics* 52(7):740–747
- Nogueira AR, Pugnana A, Ruggieri S, et al (2022) Methods and tools for causal discovery and causal inference. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 12(2):e1449
- Penedo M, Rivera-Baltanás T, Pérez-Rodríguez D, et al (2021) The role of dopamine receptors in lymphocytes and their changes in schizophrenia. *Brain, Behavior, & Immunity-Health* 12:100199
- Pingault JB, O’reilly PF, Schoeler T, et al (2018) Using genetic data to strengthen causal inference in observational research. *Nature Reviews Genetics* 19(9):566–580
- Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. *The American Journal of Human Genetics* 69(1):1–14
- Reay WR, Kiltschewskij DJ, Geaghan MP, et al (2022) Genetic estimates of correlation and causality between blood-based biomarkers and psychiatric disorders. *Science Advances* 8(14):eabj8969
- Sanderson E, Glymour MM, Holmes MV, et al (2022) Mendelian randomization. *Nature Reviews Methods Primers* 2(1):1–21
- Schuster NA, Twisk JWR, ter Riet G, et al (2021) Noncollapsibility and its role in quantifying confounding bias in logistic regression. *BMC Medical Research Methodology* 21(1). <https://doi.org/10.1186/s12874-021-01316-8>
- Sun J, Lyu R, Deng L, et al (2022) SMetABF: A rapid algorithm for Bayesian GWAS meta-analysis with a large number of studies included. *PLoS Computational Biology* 18(3):e1009948
- Tam V, Patel N, Turcotte M, et al (2019) Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics* 20(8):467–484
- Tsuda K (2020) Red blood cell abnormalities and hypertension. *Hypertension Research* 43(1):72–73
- Van Erp S, Oberski DL, Mulder J (2019) Shrinkage priors for Bayesian penalized regression. *Journal of Mathematical Psychology* 89:31–50
- Watanabe K, Taskesen E, Van Bochoven A, et al (2017) Functional mapping and annotation of genetic associations with FUMA. *Nature Communications* 8(1):1826

- Weinstein JJ, Chohan MO, Slifstein M, et al (2017) Pathway-specific dopamine abnormalities in schizophrenia. *Biological Psychiatry* 81(1):31–42
- Windmeijer F, Farbmacher H, Davies N, et al (2019) On the use of the lasso for instrumental variables estimation with some invalid instruments. *Journal of the American Statistical Association* 114(527):1339–1350
- Windmeijer F, Liang X, Hartwig FP, et al (2021) The confidence interval method for selecting valid instrumental variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 83(4):752–776
- Yang J, Ferreira T, Morris AP, et al (2012) Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nature Genetics* 44(4):369–375
- Zhang L, Pan Q, Wang Y, et al (2017) Bayesian network construction and genotype-phenotype inference using gwas statistics. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 16(2):475–489